DOCUMENT RESUME

ED 089 245 CS 001 020

AUTHOR Van Rooy, Lois
TITLE Readability Studies and the Writer of Instructional Materials.
INSTITUTION Illinois State Board of Vocational Education and Rehabilitation, Springfield. Div. of Vocational and Technical Education.; Southern Illinois Univ., Carbondale. Career Development for Children Project.
PUB DATE 73
NOTE 63p.

EDRS PRICE MF-$0.75 HC-$3.15 PLUS POSTAGE
DESCRIPTORS *Cloze Procedure; *Curriculum Development; *Readability; *Reading; Reading Achievement; Reading Comprehension; Reading Level; Reading Materials; Reading Research; *Sentence Structure; Vocabulary
IDENTIFIERS *Career Development for Children Project

ABSTRACT

The purposes of this paper are (1) to assist Career Development for Children Project (CDCP) writers in developing more readable curriculum materials and (2) to better identify research and development priorities. The first part surveys two contrasting methods that have been found useful in predicting the readability of written texts. The second part focuses on various strategies that have been found useful in controlling the linguistic variables that affect readability. The contents include: "Predicting Readability," which discusses readability formulas, guidelines for using readability formulas, the cloze procedure, concurrent validity of cloze, cloze as a validity criterion for other measurement, the cloze test and control of readability, grade level interpretation of cloze scores, and comparisons and recommendations of readability formulas and the cloze procedure; "Controlling Readability," which discusses control through word choice, sentence structure, and organization of paragraphs; "Summary"; and "References," which includes items selected from educational journals, texts, and the Educational Resources Information Center (ERIC) system. (WR)

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED 089245

READABILITY STUDIES AND THE WRITER OF INSTRUCTIONAL MATERIALS

written by

Lois Van Rooy, Consultant

for

Career Development for Children Project
Department of Occupational Education
Southern Illinois University at Carbondale
Larry J. Bailey, Project Director

in cooperation with

Research Development Unit
Division of Vocational and Technical Education
Board of Vocational Education and Rehabilitation
State of Illinois

AS 001 030

FORWARD

The Career Development for Children Project (CDCP) is a research-curriculum development project conceived primarily to develop experimental curriculum materials for grades one through eight. In pursuit of its major goal, a climate for systematic inquiry has been created in which project staff continuously analyze the technology of curriculum development, explore alternative courses of action, rigorously evaluate the products of their efforts, and often, challenge established institutions of education. One such institution--readability formulas--has long stimulated the inquisitiveness of CDCP researchers.

It is obvious to the curriculum writer and illustrator that an individual's ability to "read" text materials is dependent upon many factors. Readability measures, however, seldom take into account anything other than linguistic characteristics. A further limitation of such formulas is that they cannot be used as guidelines for writing text materials.

To assist CDCP writers in developing more "readable" curriculum materials and to better identify research and development priorities, the Director contracted with Lois Van Rooy to review and synthesize the literature related to the issues of controlling and predicting readability. The paper which follows is the result of her scholarly efforts. The first part surveys two contrasting methods that have been found useful in predicting the readability of written texts. The second part focuses on various strategies that have been found useful in controlling the linguistic variables that affect readability. The paper merits careful consideration by all who are engaged in the development of experimental educational materials.

Larry J. Bailey
Project Director

TABLE OF CONTENTS

Forward

PART 1. PREDICTING READABILITY 2

Readability Formulas 2

Guidelines for Using These Formulas 6

The Cloze Procedure 12

Concurrent Validity of the Cloze 14

The Cloze as Validity Criterion for Other Measurements 15

The Cloze Test and the Control of Readability 19

Grade Level Interpretation of Cloze Scores 20

Readability Formulas and the Cloze Procedure: Comparisons and Recommendations 23

Ease of Application 23

Accuracy of Measurement 24

PART 2. CONTROLLING READABILITY 28

Control Through Word Choice 29

Frequency 30

Word Length 31

Abstractness 32

Control Through Sentence Structure 39

Grammatical Transformation and Sentence Complexity . . .39

Sentence Length and Sentence Structure 41

Position of the Modifier: The Loose vs. the Periodic Sentence 48

Control Through Organization of Paragraphs 49

Summary and Conclusion 54

References 55

Part 1. Predicting Readability

For the past fifty years publishers and editors of children's textbooks have been using, and sometimes misusing, various of the mathematical formulas available for measuring readability,[1] formulas such as those developed in the forties by Lorge, Flesch, and Dale and Chall. A basically different method is the cloze test, a measure developed in the fifties by Taylor. Though for readability researchers the cloze procedure has been a common tool, it is less well known to others directly concerned with the readability of written instructional materials. Here, then, is a description and discussion of the two kinds of measures, followed by a comparison of their merits as readability predictors.

Readability Formulas

McLaughlin (1969b, p. 640) gives this convenient definition:

> A readability formula is simply a mathematical equation derived by regression analysis. This procedure finds the equation which best expresses the relationship between two variables, which in this case are a measure of the difficulty experienced by people reading a given text, and a measure of the linguistic characteristics of that text. This formula can then be used to

---

[1] The term readability has been variously defined as (1) the typography of a work, (2) the appeal or interest value of its subject matter, and (3) the ease or difficulty the intended reader has in comprehending its meaning (Chall, 1958, pp. 4-8; Klare, 1966b, pp. 241-242). Studies in readability for the most part have dealt with the third aspect, readability as the "quality of writing that permits a reader to read and understand it readily" (Klare, 1966b, p. 241), and it is in this restricted sense, too, that the term is used here.

predict reading difficulty from the linguistic characteristics of other texts.

Until the early fifties the focus in readability research was on developing, and then validating, statistical formulas that would enable the user to predict the level of reading ease or difficulty of existing materials. The work before 1960 is extensively surveyed in two important histories of readability research, those by Chall (1958) and Klare (1963). Chall reports no fewer than 29 readability formulas appearing between 1923 and 1953, and Klare reviews 31 such studies appearing before 1960. In recent years only a few formulas have entered the lists, notably Fry's (1968) Readability Graph and McLaughlin's (1969a) SMOG Grading.

Over the years, hundreds of elements, or variables, have been explored as predictors of readability. (A brief, up-to-date summary of empirical predictors that have been used is given in Klare, [1966b] 1971, pp. 242-246.) The aim was to develop formulas that were both accurate and easy to use, and the practice was "to prefer few variables to many, and easily counted variables to hard" (Klare, 1966a, p. 120). Gradually, two variables, easily quantified, were found to have the highest predictive value: an estimate of vocabulary difficulty combined with some measure of sentence structure. These findings were substantiated by the extensive factor analytic studies of Brinton and Danielson (1958) and Stolurow and Newman (1959).

The vocabulary factor is heavily weighted in most readability formulas, including those listed in Table 1. Words are rated as easy or difficult depending on their inclusion in a basic word list. Formulas

TABLE 1

THE LANGUAGE ELEMENTS MEASURED IN SIX FREQUENTLY USED READABILITY FORMULAS

Formula	Date Published	Variables Counted
Lorge: The Lorge Readability Index	1948; revision of 1939 and 1944	Sentence length (average number of words per sentence); prepositional phrases (ratio of phrases to total number of words); vocabulary (number of words that do not appear on Dale 769 list)
Flesch: The Flesch Reading Ease Formula	1948; revision of 1943	Sentence length (average number of words per sentence); vocabulary (number of syllables per 100 words)
Dale and Chall: A Formula for Predicting Readability	1948	Sentence length (average number of words per sentence); vocabulary (number of words that do not appear on Dale 3,000 list)
Spache: The Spache Readability Formula	1966; revision of 1953	Sentence length (average number of words per sentence); vocabulary (number of words that do not appear on Stone's revision of Dale 769 list)
Fry: The Readability Graph	1968; revision of 1963 and 1964	Sentence length (average number of sentences per 100 words); vocabulary (average number of syllables per 100 words)
McLaughlin: SMOG Grading	1969b	Vocabulary (number of polysyllabic words appearing in a total sample of 30 sentences)

using such lists include those of Lorge, Dale and Chall, and Spache. Or words are rated in terms of their length--number of letters, number of syllables, number of Latin suffixes, and the like. All of the standard formulas currently in use combine an estimate of vocabulary difficulty with some measure of sentence length, structure, or complexity, length being the measure of the sentence that is taken most frequently.[2]

Chall (1958, p. 47) wrote, "Can we actually dismiss the entire problem of difficulty as one of hard or long words and long sentences?" In a 1966 study (published in 1971), Coleman illustrated the problem (pp. 184-186):

> Formula 1: With the percentage of one-syllable words as the single measure, he can predict 73.8 percent of the variance in the difficulty of a text.
>
> Formula 2: With the addition of a second element, number of sentences per hundred words, he can predict 80.6 percent of the variance, a gain of less than 7 percent. "The gain is just barely worth the cost, and adding additional predictors becomes less and less feasible economically."
>
> Formula 3: With an additional count, number of pronouns per hundred words, he gains only 1 percent additional variance.
>
> Formula 4: With the addition of a fourth element, prepositions per hundred words, once more the gain is only 1 percent.

---

[2] SMOG Grading (see Table 1), seemingly an exception, measures sentence length nonetheless. McLaughlin (1969b, pp. 640-641) compares his formula with the "traditional type" in these terms: "Obviously, you must measure word length and sentence length separately if you are going to add the two measures together. But you achieve the equivalent of multiplying the two measures if you simply count a fixed arbitrary number of sentences and then count, say, the number of syllables within those sentences."

using such lists include those of Lorge, Dale and Chall, and Spache. Or words are rated in terms of their length--number of letters, number of syllables, number of Latin suffixes, and the like. All of the standard formulas currently in use combine an estimate of vocabulary difficulty with some measure of sentence length, structure, or complexity, length being the measure of the sentence that is taken most frequently.[2]

Chall (1958, p. 47) wrote, "Can we actually dismiss the entire problem of difficulty as one of hard or long words and long sentences?" In a 1966 study (published in 1971), Coleman illustrated the problem (pp. 184-186):

> Formula 1: With the percentage of one-syllable words as the single measure, he can predict 73.8 percent of the variance in the difficulty of a text.
>
> Formula 2: With the addition of a second element, number of sentences per hundred words, he can predict 80.6 percent of the variance, a gain of less than 7 percent. "The gain is just barely worth the cost, and adding additional predictors becomes less and less feasible economically."
>
> Formula 3: With an additional count, number of pronouns per hundred words, he gains only 1 percent additional variance.
>
> Formula 4: With the addition of a fourth element, prepositions per hundred words, once more the gain is only 1 percent.

---

[2] SMOG Grading (see Table 1), seemingly an exception, measures sentence length nonetheless. McLaughlin (1969b, pp. 640-641) compares his formula with the "traditional type" in these terms: "Obviously, you must measure word length and sentence length separately if you are going to add the two measures together. But you achieve the equivalent of multiplying the two measures if you simply count a fixed arbitrary number of sentences and then count, say, the number of syllables within those sentences."

> "In brief, formula 1, which requires a single count, predicts 73.8 percent of the variance, while the considerably more tedious counts of formula 4 predict 82.8 percent, a gain of only 9 percent."

As Chall (1958, p. 54) aptly observed in a similar context, "The law of diminishing returns sets in early in readability prediction."

Given a formula which could measure reading material on a continuum from "easy" reading to "difficult," there was needed a corresponding set of grade level interpretations. How were these grade levels determined? Authors turned to such authorities as librarians, teachers, and publishers. See Table 2.[3] They used more objective criteria: correlations with passages already graded in difficulty, such as the McCall-Crabbs Standard Test Lessons in Reading; correlations with the comprehension test scores of children and adults with known reading ability; correlations with other formulas of grade and rank assigned to the same material, and so on. Cross-validation studies, among the various formulas and between formula scores and independent measures (such as tested comprehension), are often cited in the initial presentation of the readability formulas (see Table 2). Additional studies are reviewed, for example, in Chall (1958, Chs. IV-VI).

Guidelines for Using These Formulas

In the literature authors frequently have tried to ward off various misunderstandings and misapplications. In particular they have drawn

---

[3] For a more detailed summary of various features of most of the readability formulas published between 1923 and 1953, see Tables II-IV in Chall (1958, pp. 36-39, 42-44, 48-53).

TABLE 2
THE CRITERIA UPON WHICH SIX FREQUENTLY USED READABILITY FORMULAS WERE BASED

Formula	Range of Difficulty	Material Used in Computing Formula	Criterion for Establishing Grade Levels[a]	Author's Description of Criterion
Lorge (1948)	Grades 3 to 12	McCall-Crabbs (1926)--the 376 passages in Books II to V	List of graded reading comprehension tests	"The school grade at which the passage can be understood"
Flesch (1948)	Grades 3 to 12	Selected adult magazines; McCall-Crabbs (1926)--the 376 passages in Books II to V	Flesch's judgments of various adult magazines; list of graded reading comprehension tests	"The average student will find [the reading] easy; he will be able to learn something from it"
Dale and Chall (1948)	Grades 3 to 12 (to grade 16 for passages on health)	McCall-Crabbs (1926)--the 376 passages in Books II to V; selected writings in health and social studies	List of graded reading comprehension tests; judgments of experts; reading grades of children and adults	"The grade at which a book or article can be read with understanding"
Spache (1966b)	Grades 1.5 to 3.9	224 samples from a total of 152 primary-level school books--129 basal readers and 23 social studies, health, and science texts	Publishers' designations; pupils' oral reading errors	"A child with that level of reading ability [can] read the book with adequate comprehension"
Fry (1968)	Grade 1 to college level	"Lots of books which publishers said were 3rd grade readers, 5th grade readers, etc."	Publishers' designations; pupils' comprehension test scores	[None expressed]
McLaughlin (1969b)	Grades 6 to 19 and above	McCall-Crabbs (1961)--390 passages	List of graded reading comprehension tests; pupils' tested reading efficiency	"The grades are supposed to be those which a reader needs to ensure complete comprehension"

[a] Only the criteria are listed that were cited in the original presentation of these formulas. Additionally, some of these formulas were validated initially by correlation with other formulas.

attention to the following:

Matching formulas and reading materials.--The passages being evaluated must be within a prescribed range of difficulty (see Table 2). For example, the Spache Readability Formula should not be used in estimating the reading ease of intermediate texts.

The passages should be similar to the reading material originally used in computing the formula (see Table 2). For example, in a recent investigation Froese (1971) applied the Dale-Chall formula to unlike subject matter, sixth grade science textbook materials. His criterion was the cloze procedure, which according to its originator (Taylor, 1953) and others is applicable to an unlimited range of materials. The author concluded that "the Dale-Chall formula scores for elementary science textbook material should be used cautiously."

In his report Froese made reference to an investigation by W.R. Brown (1965):

> He found that students at the seventh and eighth grade levels apparently comprehended a N.S.T.A. publication entitled Spacecraft which was rated at 11-12 grade-level by the Dale-Chall formula. Furthermore, when the vocabulary from a 1961 edition of a third-grade science textbook was accepted as familiar, it was found that it lowered the readability of the same sample from 11-12 grade-level to 9-10 grade-level. His conclusion was that when the Dale-Chall formula was applied to science textbooks it seemed to place them higher than was warranted.

Probably the most striking demonstration, and certainly the most frequently cited one, was an experiment by Taylor (1953) in which the Flesch and the Dale-Chall formulas rated Erskine Caldwell as difficult and Gertrude Stein as very easy--indeed the Dale-Chall formula rated Stein as "within the comprehension level of fourth or fifth grade

school children!"[4]

Interpreting the scores.--Referring to Chall's (1958) study, Spache (1966a, p. 35) gives this useful summary: "The estimates of difficulty are not exactly accurate but in most cases have a probable error of about one full grade. The grade level designations derived from their use will be within 8-10 months of the actual difficulty in half the estimates. The remainder of the estimates may well be in error by more than this amount."

Authors warn, too, of "the lack of one-to-one relationship between the indexes from the various formulas" (Chall, 1958, p. 169). Explanations for such discrepancies, of course, vary with the context. Here are a few examples: One formula assigned a higher grade level to a sample than a second one had; the first formula counts proper names and dates as "hard" and the second one rates them as "familiar." One formula developed recently assigned a lower grade level to a sample than an older formula had; the tentative explanation was that students today can read better than students of former years. One formula typically assigns scores two grades higher than a second one does. The criterion used in establishing grade levels for the first formula was "complete comprehension" of what was read; for the second one the criterion was less stringent: simply, "understanding" a book or an article (see Table 2).

---

[4] It is probably more than coincidence that all three studies investigated the Dale-Chall formula. If any of the formulas is generally recognized as more accurate than the others, Dale-Chall has this distinction. And see a direct comparison of the Dale-Chall and the Flesch formulas in a recalculation of those measures by Powers, Sumner, and Kearl (1958).

"Writing for a readability formula."--Of course, the problem of measuring what has already been written should be distinguished from the problem of writing what will prove readable. But, misled and confused, some writers and publishers have tried (in Lorge's words) to inflate various of the formulas into recipes for writing. Lorge (1949, p. 93) tells this story about a misuse of his formula (see Table 1):

> One person, advocating the formula as a rule for writing recommended that the sentence "I am going to town" should be rewritten as "I am going townwards." She explained that this would reduce sentence length, involve fewer different words, and eliminate one prepositional phrase. Thus, she explained, the sentence would become easier to read."

As Lorge (1949, pp. 93-94), Dale and Chall (1948, pp. 9-10), Spache (1966a, pp. 33-34), and Fry (1969, pp. 535-536), among others, have emphatically stated, readability formulas cannot be used as rules for writing.[5] Writers are advised to use the readability formulas only after a passage has been written. In any event, as Schlesinger (1968, p. 22) wrote,

> It is doubtful whether [a readability formula] can successfully estimate the readability of texts written according to the formula . . . . This is so, because one can very well "beat the formula," intentionally or not, and remove the symptom of reading difficulty without removing the under-

---

[5] Flesch takes exception to this view in "A Dissenting Opinion on Readability" (1949b, p. 334): "[Such] statements, it seems to me, dodge the main issue of readability. If readability measurement is worth anything at all, then the formulas must be usable as tools in preparing readable materials." But the only evidence he gives is the popularity of one of his formulas, used as a writing tool by "many thousands of newspapermen, advertising copywriters, textbook authors, business writers, etc."

> lying cause. For instance, when sentences of a given text are shortened, the formulas, which include sentence length as a measure, will decree that the text is now more readable, but it remains to be seen whether this is actually so.

Up until the last ten or fifteen years, most readability researchers were primarily concerned with measuring materials already written. When they addressed the writer, their suggestions invariably were the kind that have long been part of the untested advice given in traditional English composition classrooms. Recently, however, linguists and psychologists have taken an accelerated interest in the problems of the writer of instructional materials, and empirically based assistance is at hand. See Part 2 of this paper.

As has been noted, these formulas predict readability through measurement of a limited number of elements, factors which are at once easily counted and significantly correlated with reading ease or difficulty. But they ignore other, less easily quantified aspects known to be related to comprehensibility. As Dale and Chall wrote (1949, p. 25), in summing up the art of readability measurement as realized over twenty years ago,

> On the problem of comprehensibility, we have made some strides. The numerous readability formulae help give a rough approximation of the difficulty of a piece of material. Some of these are very easy to apply--and consist of mechanical counting of words, syllables, length of sentences, prepositional phrases, etc. But because they are mechanical, they are usually taken as infallible. The important factors of conceptual difficulty, organization of the material or the logic, semantic variations in words, etc. have been discussed widely in the literature in readability, but have not yet been incorporated in any formula.
>
> To date, there is no composite method that can be used to measure all aspects of readability. We must consider separately the aspects of format and organization, content, expressional elements; and then make a judgment as to the

suitability of a particular book for a particular group.

> Some day, though, we may be able to say that a given piece of material is readable for a particular group of readers and have this statement encompass all the possible factors that contribute to its readability. At this present time, however, we can say only that it is readable on the basis of such and such a criterion taking such and such factors into consideration.

Ironically, four years after this appraisal a method was developed that seemingly meets all of Dale and Chall's criteria except that it is not a formula. It is called the cloze test.

The Cloze Procedure

Taylor (1953) originated the cloze, an objective, easily applied method for studying communication which appeared to be a measure of the combined influences of all variables that affect comprehensibility (Taylor, 1953, p. 432). It is not a readability formula at all. Where formulas count elements in a passage, the cloze procedure counts correct responses by a reader (or listener or viewer). Where readability formulas measure the stimulus, cloze scores measure the response.[6]

Following is the method as outlined by Bormuth (1968, p. 429):

> A somewhat oversimplified description of the cloze readability procedure includes these steps: (a) passages are selected from the material whose difficulty is being

---

[6] As a measuring instrument the cloze procedure has been used not only to predict readability of passages but also to evaluate individual language aptitudes, reading ability, information gain, and so on. Examination of the cloze test other than as a readability measure is beyond the scope of this paper. For extensive analyses of a variety of applications, see Rankin (1965) and Jongsma (1971).

> evaluated, (b) every fifth word in the passages is deleted and replaced by underlined blanks of a standard length, (c) the tests are duplicated and given, without time limits, to students who have not read the passages from which the tests were made, (d) the students are instructed to write in each blank the word they think was deleted, (e) responses are scored correct when they exactly match (disregarding minor misspellings) the words deleted. When the tests have been made properly, a student's score can be interpreted as a measure of how well he understands the materials from which the tests were made.

The set of procedures described by Bormuth is the most frequently used one. Alternative forms of the test have been constructed. For example, deletions have been made at intervals greater than five words; the tests have been given to students who had previously read the "unmutilated" passages; credit has sometimes been given for synonyms as well as exact responses. These alternatives are neither as economical nor as convenient as the particular set of procedures described by Bormuth, above, and using the simpler set does not materially affect the validity of the test. These and other variations in cloze test construction are discussed in full in the literature.[7] See Rankin (1965) and Jongsma (1971), who have themselves worked with this measure, for extensive analyses of methodological considerations in using the cloze technique.

---

[7] A variation investigated by Anderson (1970) is of interest. He compared the effect of using underlined blanks of a standard length with the effect of using blanks the same length as the deleted words and found both versions equally valid. "The implication of these findings is that by the use of a photocopy procedure, such factors as size of print, illustrative material, and page layout may be included in the estimate of the ease or difficulty of printed material as measured by cloze tests." His findings require validation, it would seem, on a population closer to home, for his subjects were "indigenous primary-school pupils in New Guinea for whom English was a foreign language."

Concurrent Validity of the Cloze

What, exactly, do cloze tests measure? Are they predictors of what is commonly regarded as comprehensibility?

Comprehension has traditionally been measured as the ability to answer questions about a passage (MacGinitie, [1966] 1971, p. 207). Numerous studies of correlations between cloze tests and other reading tests covering the same passages have revealed that the cloze procedure has considerable concurrent validity. See Rankin (1965) and Jongsma (1971). But in a factor analytic study, in which they compared results achieved from cloze tests and a number of standardized tests of reading, listening, and language-symbolizing ability, Weaver and Kingston (1963) found low correlations in respect to "verbal comprehension." Rather, they found that the abilities used to complete cloze deletions were related to "redundancy utilization."

In an informal sense, redundancy may be defined as "using more than the 'minimum' number of words needed to get an idea across"; it is the opposite of "terseness" (Hafner, 1965, pp. 152-153). Referring to the Weaver and Kingston (1963) study, MacGinitie ([1966] 1971, pp. 208-209) pointed out that one does not need to comprehend what he is reading in order to fill in the blanks; "language is highly redundant, and subjects can often restore words successfully with only a recognition of familiar patterns of expression and no real understanding." Instead of the cloze procedure, which is "no more valid," MacGinitie ([1966] 1971, pp. 206-207) suggested going back to the less reliable "well-devised set of questions" as the criterion of comprehensibility.

> Ability to answer questions about the information in the passage has been the traditional measure of understanding. This measure at least has good face validity. We could do a lot toward standardizing and specifying the nature of the questions that are asked.

As Rankin (1965) pointed out in his survey, of the numerous studies that he reviewed only the Weaver and Kingston (1963) investigation found low correlations between cloze tests and standardized reading tests; only they found that cloze tests had low loadings on "verbal comprehension." Their view of the cloze as something less than a measure of comprehension places them clearly in the minority, then. Other investigators, for example those who have evaluated the cloze chiefly as a readability measure, have obtained high correlations between the rankings assigned by cloze scores and by multiple choice tests made from the same passages (e.g., Taylor, 1953: a correlation of .76; Bormuth, 1963: a correlation of .92).

Rankin concluded his extensive analysis of the cloze (1965, pp. 147-148) in these terms:

> It may be said that there is abundant evidence pointing to the validity and usefulness of this technique as a measuring instrument. Correlations with various validity criteria are sufficiently high to show that it has satisfactory concurrent validity . . . .

The Cloze as Validity Criterion for Other Measurements

In his original presentation of the cloze, Taylor (1953) compared rankings assigned by readability formulas with those obtained by the cloze test. He found that the cloze procedure could rank relatively easily worded passages of high concept load, such as those written by Gertrude Stein and James Joyce, better than the Flesch and the Dale-

Chall formulas could (see pp. 8-9 of this study).

Since 1953 the cloze test has increasingly been used as a criterion for evaluating readability formulas. Fry's application is typical: in validating his Readability Graph at primary levels, Fry (1969) compared the rankings assigned by his method with those obtained using the Spache Readability Formula, the cloze procedure, and oral reading errors. He found that "The cloze method seemed to be the most accurate and made the finest distinctions [p. 536]."[8] (Note, too, the reference to the Froese [1971] correlational study on p. 8 of this paper.)

Other applications of the cloze have been considerably broader in scope and implication, for controlling as well as for predicting the readability of language. A brief review of two such contributions serves to reveal the range of current readability research, advances which may be attributed in large part to the development of the cloze technique with its greater flexibility: it can be used to measure the difficulty not only of whole passages but also of units as small as a word. The studies to be reviewed are Bormuth's controversial "Readability: A New Approach" (1966) and the Miller-Coleman Readability Scale (Miller & Coleman, 1967).

---

[8] Fry gave the following reasons for not preferring the cloze test, with its greater validity, over his Readability Graph (1969, p. 536): "Were it not for the enormous amount of time this method takes, cloze procedure would be an excellent way to determine readability. In addition to the time it takes to make the cloze passages, a number of different passages must be tested at the same time on the same group of children. One cannot return to the same group of children several months later, for their reading abilities will have changed and the cloze error scores will not be comparable. As a research tool the method is excellent but for practical purposes it is all but impossible to use."

Bormuth (1966) viewed the cloze test as "the crucial factor" in advancing readability research, for now investigators had a tool that was at the same time at least as valid as and more reliable than the best practice available in the past: giving subjects multiple choice tests over a passage (see the Lorge, Flesch, Dale-Chall, and McLaughlin entries in Table 2, p.7). Additionally, the cloze was more flexible: it could measure units as small as an individual word or sentence.

Using the cloze procedure as the criterion, Bormuth investigated five problems which he considered basic to the development of accurate formulas (quoted from Klare, 1966a, p. 119):

1. Is there a linear relationship between linguistic variables and comprehension difficulty?
2. Do linguistic variables have the same influence upon difficulty for readers of different levels of achievement?
3. Can the readability of such small language units as individual words and sentences be measured?
4. Can the validities of present readability formulas be improved?
5. Can new linguistic (or psycholinguistic) variables be used profitably in readability research?

Among the five, Bormuth considered the most important question raised in his study to be whether present readability formulas can be improved. He thought they could. As he wrote elsewhere (1967, p. 845),

> The reason lies largely in the fact that researchers in several disciplines have developed research tools which have aided greatly the study of readability. Psychologists have developed the cloze procedure into an accurate and reliable method of measuring language difficulty. Linguists have developed descriptions of various features of language and these descriptive devices have been further adapted into powerful new techniques for measuring the features of language that influence its comprehension

> difficulty. Finally, advances in our understanding of the mathematics used in our analyses have led to improved designs for readability formulas.

In his extensive investigations Bormuth (1966) used a number of improved analytic techniques, and he measured the predictive value of a great many new linguistic variables, some of which had not before been incorporated in a readability formula. He found several with high correlations with passage difficulty. For example, the ratio between the number of pronouns and the number of conjunctions in a passage correlated .81 with difficulty. Working with a combination of variables, he attained a multiple correlation of .93 with passage difficulty.

Had Bormuth achieved the most important of his goals, to demonstrate that the accuracy of readability formulas can be vastly improved? He wrote (p.109),

> In the past readability formulas have received widespread use and have had validities ranging from .5 to .7. If these validities may be taken as a criterion of usefulness, then the formulas represented by the multiple correlations in this analysis must be said to be useful.

Will general readability formulas in the future attain validity correlations approaching 1.0? Bormuth thinks that they will; he foresees the improvement to be chiefly a consequence of the greater validity of new linguistic variables (1966, p. 128). But Coleman ([1966] 1971), Klare (1966a), and MacGinitie and Tretiak (1971) have their doubts. For a convenient summary and explanation of opposing views, see Klare (1966a). The question has not, so far, been answered; it must be referred to the researcher for additional study. In the meanwhile the cloze procedure is recommended for general use; detailed instructions, including practical suggestions for administering the

test and analyzing the results, have been published in the Elementary English Journal (Bormuth, 1968).

A second valuable contribution to current research in readability is the Miller-Coleman Readability Scale, or MCRS (Miller & Coleman, 1967). It, too, used the cloze as its criterion. It is a set of 36 150-word passages taken mostly from the McCall-Crabbs Standard Test Lessons in Reading (1926); its range in difficulty is "from a first-grade reader to very difficult technical prose [Miller & Coleman, 1967, p. 851]."

The passages were originally calibrated using three different types of cloze tests. Subsequently Coleman and Miller (1968) calibrated the scale using a new index, Information Gain (IG), which measures the efficiency with which a passage transmits new information. (IG is, simply, the cloze score obtained on a first reading subtracted from the cloze score obtained on a second reading.) Later Aquino (1969) validated the MCRS using two additional measures, word-for-word recall and subjective judgment. He obtained correlations with cloze scores that ranged from .89 to .94. As he pointed out, his study was an investigation of both the validity of the cloze technique and the precision of the MCRS. He reconfirmed the suitability of the cloze test as a measure of passage difficulty (p. 347). And he found the MCRS to be "a precisely graded scale useful for preliminary correlation studies or for computing readability formulas [p. 347]."

The Cloze Test and the Control of Readability

We have been taking an exceedingly brief look at only a few of the applications of the cloze test. It has been used both as an end

in itself--to determine the readability of specific passages--and as a means to an end. It has proven itself an important tool in the study of language. For example, the cloze test has been used, most extensively by Bormuth (1966) and Coleman ([1966] 1971), to seek out new linguistic variables not only for their potential value as indexes of readability, to be incorporated eventually into a formula, but also for the guidelines that these variables might furnish the writer who would control the readability level of what he writes. See Part 2 of this paper.

Grade Level Interpretation of Cloze Scores

As Rankin noted (1970, p. 2),

> For some time the main weakness of the cloze procedure as a measure of readability was the absence of criteria for interpreting raw scores. The relative difficulty of two or more passages could be determined, but no interpretation could be placed upon the difficulty of each passage.

Bormuth has led in the work of establishing criterion reference scores. In the first of his investigations he related cloze indexes to conventional multiple choice comprehension test scores. As he explained the procedure (1968, p. 433),

> A standard has long been accepted for conventional comprehension tests and this standard is widely used in practice. It asserts that materials are suitable for use in a child's instruction when he is able to answer correctly 75 percent of the questions asked him about the materials. The materials are said to be suitable for his independent study when he can answer 90 percent of the items. Bormuth found that a score of 75 percent on conventional comprehension tests is comparable to a score of 44 percent on a cloze readability test made from the same passage and that answering 57 percent of the cloze items is comparable to answering 90 percent of the items on conventional comprehension tests.

Rankin (1970) has since corroborated Bormuth's findings.

Bormuth's second approach to the study of criterion levels of performance differed significantly from the first. In it Bormuth (1971) related cloze scores not to the conventional standards used previously but to such bases as preference ratings that the student himself expressed in interest inventories, classifying passages as to whether he felt them appropriate for use in a textbook, a reference work, or voluntary reading material. As outlined in the abstract,

> The purpose of these studies was to develop and demonstrate a model for identifying criterion levels of performance that can be rationally defended as being the best level of performance for a particular instructional task. The specific objective was to identify the score on a cloze test that represents the most desirable level of performance on instructional materials.

The subjects were hundreds of pupils in grades 3 to 12 in a number of Chicago suburban schools.

> The studies were designed to permit the results to be generalized to students in grades 3 through 12, to materials on most of the topics and at most of the difficulty levels that these students are likely to encounter in instruction, and to each of the major purposes, for which the students are likely to read a passage. [p. ix]

In all, six measures were weighted in the study: information gain, rate of reading, willingness to study, and preferences for the subject matter, style, and level of difficulty (p. ix). Teachers themselves weighted each of the measures, "an expedient rationalized by the suppositions that teachers are acquainted with the goals of instruction, the values society places on those goals, and how the variables in the model contribute to the attainment of those goals [p. ix]."

The criterion scores presented in Bormuth's model differ markedly from those obtained when the comprehension test scores traditionally used were accepted as a standard (see p. 20 of this paper). First, Bormuth's model required that a single scale for all grade levels be replaced by a scale identifying a separate set of scores for each grade level. Second, and unexpectedly, it reversed the assumption that material designed for individual use should be easier to read than material intended for group study; it stipulated (at least for grades 5 through 12) that

> Students should receive easier materials for use in their supervised study, for textbook reading, than they should get for their independent study, for reference and voluntary reading. What makes this a result of major interest is the fact that it runs exactly counter to the practices recommended by teacher trainers in the area of reading. [p. 142]

Bormuth sees the chief value of his (1971) study not so much in the particular scores obtained--they are not yet ready to be put into the hands of educators--but in the methodology itself:

> The major strength in the criterion scores presented here lies in the model and the data by which they were derived. That is, a criterion score derives its validity from what it produces for its user. The traditional criterion scores produce essentially unknown outcomes for their users, whereas the present criterion scores are supplied with a considerable amount of data to describe many of their effects. However, it must be emphasized that this is not to claim that the criterion scores presented here were good in any absolute sense. Rather, it is claimed here that they are far better than the criterion scores presently in use. The earlier discussions in this report should have made it abundantly clear that much work remains before criterion scores that are good in an absolute sense can be supplied to educators. The identification of performance criterion scores for criterion reference tests is analogous in many ways to assigning norms to norm reference tests. And there is no reason to believe that developing the theory and technology for identifying

> performance criterion scores will be any less complex or less energy-consuming than the development of the theory and technology for assigning norms. [p. 148]

Bormuth's model represents a turn-about in its approach to the problem of identifying suitable instructional materials. Where before a student's score was compared with a previously established norm, the average of comprehension test scores attained by a representative sample, now his score could be compared with a standard to be met if that student were to achieve a "known outcome," a specified educational objective. Once the cloze scores identified in Bormuth's model have been refined, both for individual grade levels and for the particular use to be made of the materials, then the cloze procedure indeed will have become a powerful differentiating tool, one able to predict with a fair amount of certainty the success that a student will have with a given set of materials.

Readability Formulas and the Cloze Procedure: Comparisons and Recommendations

Klare (1966a) has summarized the conflicting demands made of a measure of readability: simultaneously it must be easy to use and accurate in its predictions.

Ease of Application

Certainly any of the numerous conventional formulas is a more convenient and economical measure than the cloze procedure. The originator of the cloze test (Taylor, 1953) admitted that

> [The Flesch, Dale-Chall,] and most other formulas . . . . are easier and quicker to apply. Their use does not require word deletion, the reproduction of materials,

> experimental controls, and representative population samples. [p. 433]

In Fry's opinion, "For all practical purposes [the cloze procedure] is all but impossible to use" (see footnote on p. 16 of this paper). Others have minimized the time and inconvenience; for example, Hafner (1965) has recommended that "Cloze procedures . . . be used by publishers and authors as a fairly inexpensive method of conducting field trials of the readability of early drafts of chapters in their books." [p. 153]

Whether the cloze procedure is judged to be practical would seem to depend in part on whether the material to be tested is "standard" (see pp. 8-9 of this paper) and whether the intended audience is "average," for, as Taylor has pointed out (1953), readability formulas do not necessarily apply to particular populations:

> It is a little unreasonable that a single readability score for an article on cattle breeding should apply alike to residents of Texas "cow country" and metropolitan Brooklyn. In such cases, it appears that the user of a formula might employ cloze procedures to check up on his results. [p. 433]

Even more important because many more readers stand to benefit, the decision to use the cloze procedure depends on how widely the material is to be distributed (whether copies are to be printed by the tens or by the thousands, as with textbooks). Maturity, motivation, and other characteristics of the intended audience are additional determinants.

Accuracy of Measurement

If conventional readability formulas are easier and less expensive to apply than the cloze procedure, there seems little doubt that they are also less accurate, less able to make the fine distinctions possible with the cloze procedure, as for example Fry has pointed out (see p. 16 of this paper). Taylor (1953) considered readability formulas accurate

enough for some uses:

> For what may be called "standard" materials, these formulas seem reasonably accurate--the occurrences of the elements they choose to count usually do correlate better than chance with such criteria of validity as comprehension test scores and lists of graded readings.
>
> And they are "reliable." With relatively little training, different users of the same formula get virtually identical results for the same materials. Also, the results of different formulas have often been shown to correlate significantly. [p. 433]

In contrast, Bormuth (1966) has taken the position that

> It is problematic whether presently available readability formulas help more than they hinder. Because these formulas are easy and inexpensive to apply, they enjoy widespread use by publishers and educators. Publishers use them for "adjusting" the difficulty of instructional materials, and educators use them to decide if instructional materials are suitable for students at a given level of reading ability. Chall (1958) has made a strong case that the formulas are not sufficiently accurate to warrant either of these uses. Their validity correlations range from .5 to only .7, and experiments have shown that they have little, if any, validity when they are used as style guides for "adjusting" the difficulty of materials. Hence, the publishers' "adjustments" of the materials probably do not have the desired effect on the actual difficulty of the materials. But the practice does mislead educators. Since educators use essentially the same formulas as the publishers, they believe that the materials are suitable for their students when, in fact, they are not. [pp. 81-82]

Lessening their accuracy, readability formulas in time become obsolete, as a result of their having been validated on criteria established in the past. For example, Fry (1968) has pointed to the change in reading ability of populations as one possible reason for the lower scores attained by his Readability Graph. In reporting the results of a comparison with another, older formula he noted,

> The Dale-Chall ranks several books a little harder than the Readability Graph but perhaps the fact that the Dale-Chall was developed about 20 years ago accounts for this.

> At least it is hopeful to think that present 6th and 9th graders can read a little better. [p. 516]

Stronger support for the assumption, that "a 9th grade student today reads better than a 9th grade student in former years," might be found in recalculations of older readability formulas, using more recent criteria for establishing grade levels. As noted in Table 2, the 1926 edition of the McCall-Crabbs Standard Test Lessons in Reading provided the criterion for the Lorge, Flesch, and Dale-Chall formulas. In revising the Lorge formula, MacGinitie and Tretiak (1971) used the grade levels for the 1961 edition of the Test Lessons. Their rationale was that "The vocabulary and sentence structures that are commonly used, and that children at a given grade level can typically understand, change with time [p. 367]." However, they did not include a direct comparison of the original and the recalculated formula, such as one given by Powers, Sumner, and Kearl (1958). In revising the Flesch and Dale-Chall formulas, this group of investigators substituted the 1950 for the 1926 edition of the Test Lessons. Following is an example of their findings:

> To assess the practical significance of the revision, the original and recalculated forms of the Flesch and Dale-Chall formulas were applied to 47 sample passages from a variety of sources. The recalculated Dale-Chall formula consistently gave lower scores than the original; the average discrepancy (average absolute deviation) between the two was .94 grades. The average discrepancy between the original and the recalculated Flesch formulas was .85 grades, with the recalculated formula giving a lower score about two-thirds of the time. [p. 101]

The reading ability of a population undoubtedly changes over the years; additionally the words it uses vary. Today, television in particular is almost certainly contributing to vast "language

development mutations" that have potential impact on a population's ability to read--the words it knows, the sentence structures it is familiar with, and so on. If such changes--in the language itself or in people's ability to use it--affect the validity of particular readability formulas, they have no such effect on the cloze procedure, for it does not measure the elements of communication alone, taken in isolation, but people's response to those elements.

Rankin (1970) has conveniently summed up the advantages of the cloze test over presently available readability formulas:

> In contrast to readability formulae, the cloze procedure measures readability directly from a person, not from material alone. It measures specific contemporary target groups, and thus it need not rely on norms established in the past. It measures readability in relation to a given background of experience and in a motivational context. Unlike many formulae, it is not fooled by long, easy sentences or short, hard words. It assesses language correspondence between the author and the target group. It reflects the redundancy in the passage. In so far as factors like organization and subtle elements of style influence comprehension, this is reflected in a cloze score. It has even been found by Rankin and Culhane (1970) that pictures may influence cloze scores. Furthermore, cloze tests are easily constructed and scored . . . . [p. 1]

To this day, perhaps for some investigators the ultimate objective of readability research remains to find a measure that is at once as convenient and economical to apply as the formulas in present use and as accurate in its estimates as the cloze test. Which of the two kinds of predictors a person chooses depends on how far-reaching the decisions are that will be based on the predicted outcome. Additionally it depends on how "typical" the material and the intended audience are; for some materials, and materials in some contexts, cannot be accurately evaluated by presently available readability formulas.

Part 2. Controlling Readability

As has been seen, until the early 1950's the goal of most readability research was to discover easily counted variables that would lead to increasingly accurate formulas. The focus had been on predicting ease of reading. With the development of the cloze technique, researchers had a far more flexible tool: where most readability formulas had been designed to measure the readability of a whole passage, the cloze procedure could take the measure of a unit as small as a word. The cloze technique could, then, be used in studies focusing alike on the prediction and the control of language variables that affect reading difficulty.

Of course, the cloze test is not the only method that investigators have employed in dealing with questions about how to control comprehensibility. They have also used instruments designed to manipulate linguistic variables experimentally. In fact, Schlesinger (1968, pp. 154-155) has argued that "wherever there is a question of studying the fine-grain structure of a sentence," as for example when the order of words is under study, then methods other than the cloze ought to be used.

Studies based on a wide variety of approaches have appeared recently, which seek to test specific hypotheses about the language, such as whether simplifying verbs is equally helpful to advantaged and to disadvantaged high school students, whether splitting up complex sentences might not have adverse affects upon at least older readers, and whether paragraphs ordered inductively are easier to read than

paragraphs ordered deductively. From such studies can be derived "rules for writing" more firmly based than advice which is the outcome of whim and tradition.

The section following is a review of some of the practical suggestions that have arisen out of current linguistic and psycholinguistic research, suggestions that are of immediate, practical use to the writer of children's instructional materials. References are made to studies both in readability and in related areas, for example, research in the language development of children and the strategies for teaching reading and writing that have grown out of them. From time to time, findings of empirically based research will be augmented with rationally based strategies developed by specialists in other areas, for example English composition.

Control Through Word Choice

The vocabulary factor is heavily weighted in almost all readability formulas, and according to numerous research studies it is probably the single most important variable affecting readability. But studies have also shown that, by itself, simplifying vocabulary fails to increase comprehensibility; it must be accompanied by other changes, in sentence and paragraph structure, for example. See the extensive reviews of this research in Serra (1954, pp. 77-81) and Chall (1958, pp. 97-112).

Several interacting word factors that have been found to correlate with reading ease are frequency, word length, and abstractness. These

variables, together with implications for the writer, are described in the following section.

Frequency

Word lists are usually based on one of two counts: the frequency with which words occur in material written for a given population (or in the writing or speech of that population) and the familiarity of words, measured by testing a population on their knowledge of words presented to them. Authors who accompany their readability formulas with word lists (as opposed to counts of the number of letters, syllables, Latin suffixes, and so on) sometimes recommend that writers refer to these lists as one means of improving comprehensibility.[1] But, as a number of investigators have recently affirmed (see Chester's extensive review [1971]), these lists may be poorly founded; and they are often obsolete. A writer may wish to look for other, experimentally sound, more up-to-date lists that reflect the world of the 1970's.

Certainly if a writer is addressing very young children he will find a word list useful. If five- and six-year-olds are to be his audience he might refer to the inventory that Coleman (1970) derived from investigations of the ease with which children learned 160 common words (for the bases upon which these words were selected, see Coleman, 1970, p. 4). Subjects in his experiment were "150 preschool children between the ages of 48 and 75 months who had had no training in reading."

---

[1] Dolch (1949) has listed a number of warnings to keep in mind if one would use such lists intelligently.

If the writer is addressing older children he might find useful an up-to-date list based on the frequency with which certain words appear in the language of a given population. For all but the youngest readers, upon whom the effect appears to be lost (Chester, 1971; Coleman, 1970), common words are thought to be more comprehensible than rare words are.[2]

Or the writer can turn to a list based on a familiarity count. The most recent is Dale's as yet unpublished study, The Words We Know: A National Inventory,[3] which "includes familiarity scores in grades four, six, eight, ten, twelve, thirteen, and sixteen on about 45,000 words [Seels & Dale, 1971, p. 10]."

Word Length

A writer may distrust a vocabulary list for the same reason that he is wary of definitions found in a dictionary: language is highly mutable, and such references are out of date almost as soon as they are published. Alternatively the writer may try simply to control the length of his words, preferring one- to three-syllable words and so on.

A useful control of word length may be inferred from a study by Aquino, Mosberg, and Sharron (1968). They investigated the effect on cloze performance of material that contained a high proportion of Latin suffixes and compared the results with scores on material that had

---

[2] For an extensive review of studies investigating the relation between word frequency and comprehension, see Klare (1968).

[3] Dale has indicated in a letter that this inventory may be ready for publication in late fall 1973; in the meanwhile xerox copies of an interim report, Children's Knowledge of Words, are available at a cost of $7.50 each.

medium and low levels. The subjects were 225 eighth graders from a Southern California school district, described as "a predominantly lower-middle to upper-middle class community." The materials were 90 newspaper articles from three content areas: movies-television-theater, science, and human interest. Cloze scores on articles with a high frequency of Latin suffixes were found to be significantly lower than on material with either a medium or a low proportion of Latin suffixes. "However, the fact that cloze performance was not differentially affected by the Low and Medium levels suggests that a rather high level of L-suffixes is necessary before this variable appreciably affects cloze scores [p. 10]."

The implications of this study are clear: a writer should avoid the excessive use of words with Latin endings, such as -ism, -ment, -ation, and so on.

Aquino and others (1968, p. 10) noted that, "as a variable, L-suffix density covaries with several other variables. Words containing L-suffixes tend to be longer words than non-L-suffixed words and typically less frequent or familiar in the language." Typically, such words also give the impression of being abstract and therefore harder.

Abstractness

The terms "abstract" and "concrete" have been variously defined. A word is said to be abstract when it names an idea that cannot be perceived by the senses; conversely, concrete words name an idea that can be seen, heard, touched, smelled, or tasted. Concrete terms, then,

have "picturability." As Roger Brown (1958) qualified further,

> Of course the concrete noun, like the abstract, names a category rather than a particular instance. However, some categories have a more or less characteristic visual contour and size while others do not. Visual contour is a criterial attribute for table but not for thing or experience. [p. 266]

The sense of "concrete-abstract" that Brown preferred, "one that is congruent with more usage than is any other single sense," is defined as follows:

> A superordinate category is more abstract than its subordinate. The subordinate is more concrete. The following categories are listed in abstract to concrete order: living things, animals, vertebrates, primates, men, American men, Ralph Jones a particular American. Each of these categories is superordinate to all of those that follow it.[4]

Numerous studies in the psychology of language (psycholinguistics) have shown that the more abstract a word is, the harder it is to learn or to remember. See the reviews by Gorman (1961) and Paivio (1969). And, though the issue is by no means resolved, some studies (for example, Paivio, 1967) would appear to suggest that specificity has the same effect on ease of reading that concreteness does. In common usage the terms "abstract-concrete" and "general-specific" are interchangeable, and for our purposes there seems little point in distinguishing between these superordinate-subordinate pairs. A related pair,

---

[4] One composition specialist presumably based his "principle of generality" on Brown's definition. In How To Organize What You Write, Johnson (1964) showed how the principle might be applied in improving the comprehensibility of sentences and paragraphs as well as larger units. Christensen, too, in his extensive Rhetoric Program (1968a) included as a "first principle of writing" the recognition of levels of generality. His approach is described in a later section of this paper.

which psychologists have also examined, is "whole-part." Teachers of composition find a fourth rating useful: "plural-singular." Often two or more of these dichotomies may be seen working at the same time, as in this descriptive sentence by Kay Boyle (quoted in Christensen, 1968a, p. 43):

> The lighter's flame lighted up his features for an instant,
> the packed rosy jowl,
> the graying temple under Tyrolean hat's brim,
> the bulging, blue, glazed eye.

In relation to its subordinates (jowl, temple, and eye), the term features is simultaneously abstract, general, and plural.

Typically, the writer addressing children is scrupulous about replacing abstract with concrete nouns. Where possible he replaces youth (abstract) with a youth (concrete) and so on--examples abound in manuals of style. For his impression of children's verbal capacity matches that of Brown (1958, p. 247), who after examining the recorded speech of 24 children in a Harvard University preschool wrote, "Nouns commonly heard were truck, blocks, and teacher. The non thing-like nouns were uncommon. There were no uses of thought or virtue or attitude." Brown hypothesized that when a child uses an abstract noun he may have previously assigned to it a concrete meaning. "When the word justice comes into one's vocabulary it comes as a noun and may, as a consequence, be endowed with thing-like attributes borrowed from blocks and trucks [p. 247]." Indeed, a recent study (Lundsteen, 1971) has shown that younger readers tend to assign concrete rather than abstract meanings to words regardless of the context. Materials used were a "Choose a Meaning Test" of word meaning, a "Depth of Meaning Test" of paragraph meaning, and a

"Creative and Critical Reading Test" of paragraph meaning. Subjects were 79 third graders and 111 sixth graders in a Berkeley, California school.

The control of abstractness levels may well be more important in writing intended for children than for adults. But the rule holds also for the writer addressing more mature readers: write in terms as concrete and as specific as possible.

Not only nouns themselves but also the various word classes and subclasses may be compared on the basis of their readability, including sometimes their abstractness ratings. One recent investigation serves as an illustration. In an extensive correlational study, Coleman ([1966] 1971) examined 44 word classes, including subclasses of nouns and verbs, in order to determine which of them correlate positively with comprehensibility and which correlate negatively. The instrument used was the Miller-Coleman Readability Scale (MCRS), a set of 36 150-word passages "chosen from materials which ranged in difficulty from a first-grade reader to very difficult technical prose." The original subjects were 479 college students.

Coleman classified each one of the 5,400 words appearing in the MCRS in terms of its word class and found the average cloze score for each one of these word classes. Not unexpectedly he found that the number of concrete nouns, the number of active verbs, and the number of pronouns correlate positively with comprehensibility, that the number of prepositions and adjectives correlates negatively, and so on (for a discussion of his findings, see Coleman, [1966] 1971, pp. 165-172).

Innumerable inferences can be drawn. An extremely useful strategy recommended by Coleman simultaneously decreases the number of abstract nouns and increases the number of active verbs.[5] His advice is simple: transform nouns derived from verbs (operation, explanation, admission) to active verbs (operate, explain, admit). Replace this sentence--

> The inclusion of this man is emphasis upon the importance of the group--

with this one:

> When I included this man, I emphasized that the group was important.

As Coleman wrote (p. 169), "the finite verb is far more specific than the abstract noun. By transforming inclusion to included, tense, voice, aspect, and mood are made specific." Of course, other changes in word class are apparent in the revised sentence. The number of pronouns has been increased from zero to two; the number of prepositions has been reduced from three to zero; the copula "is" has been replaced by an active verb.

---

[5] In his study of the verb Coleman ([1966] 1971) examined the following subclasses:

> 1. V_{copula}--linking verbs such as is.
> 2. V_{full}--verbs, with the exception of copulas, that have tense markers, i.e., transitive and intransitive verbs such as hit and sleep. This category does not include the nonfinite forms (gerunds, infinitives, and participles).
> 3. $V_{nominal}$--nouns derived from verbs (explanation, knowledge). These words are not verbs at all, but it is interesting to examine them here because they can be substituted for verbs by applying a grammatical transformation. [pp. 167-168]

The investigator found an insignificant correlation between cloze scores and the number of V_{copula}, a high, positive correlation of .66 between cloze scores and the number of V_{full}, and a high, negative correlation of -.76 between cloze scores and $V_{nominal}$.

Some writing is difficult to comprehend because the subject matter itself is abstract; some "is abstract for no better reason than that the writer chose one derivative of a verb instead of another [Coleman, (1966) 1971, p. 167]." This kind of abstractness Coleman has aptly labeled "superfluous complexity." The term applies equally well to writing sprinkled with unwarranted rare, unfamiliar words and with long words where short ones would do.

In the preceding sample sentences, not only have word classes and subclasses been exchanged but also the sentence structure has been altered; for example, the main clause has been shortened from 20 to 11 syllables, from 13 words to 7. Sentence structure as an aspect of reading ease is the subject of the next section.

In summary, word difficulty has long been known to have an important effect on the readability of written material, and one of the solutions has been to construct word lists. But there is a danger here: such lists may encourage their users to think of words in isolation, as "[having] the power to operate sui generis, as if the panes in a window exist apart from the frame [Warfel, 1962, p. 141]." To take the measure of words out of their context--to view them "in the sterile and artificial atmosphere of a word list or in isolation"--is to "[ignore] a great body of linguistic and reading research [Moir, 1970, p. 217]." As Warfel affirmed,

> The three elements in a language--in descending order of importance--are the system or code, the tune, and the words . . . . The "big" words are only as important as the system makes them.[p. 141]

Some of the ways in which the system, or structure, of the language can be manipulated to improve readability are discussed in pages following, in the context of sentences and paragraphs.

Control Through Sentence Structure

The psycholinguists Coleman (1965, [1966] 1971, 1968) and Schlesinger (1968) have led in making systematic investigations of the relation between comprehensibility and the syntax of sentences, and they have challenged other researchers to take up the question. In an introduction Coleman ([1966] 1971) urged psychologists to leave their lists of unconnected words and nonsense syllables and get on with studies of the "sort of verbal behavior everyone else is interested in." Similarly, Schlesinger (1968) wrote,

> [The] neglect of research pertaining to the relationships obtaining between syntactic variables and readability is paralleled by a general neglect of these variables in psycholinguistic research. Most psychological studies on linguistic variables are concerned with words; psychologists seem to hesitate to come to grips with larger and more complex units. This is perhaps understandable in view of the great difficulties attendant on research with such units . . . ; but as long as syntactic variables are ignored, no understanding of language behavior can be achieved. [p. 16]

Some of their findings, together with the work of others, are reported in pages following.

Grammatical Transformations and Sentence Complexity

A new theory of language formulated by Chomsky (1957), a transformational-generative grammar, provided a means of evaluating sentence complexity that went beyond traditional indexes, for example, a simple count of the average number of words in a sentence. Chomsky held that to interpret a transformed sentence--Did John eat an apple? John did not eat an apple. What did John eat? An apple was eaten by John.--

people must first transform it back into its kernel--John ate an apple.

Generally, grammatical transformations have been found more difficult than their underlying kernels. For a "state of the art" report of studies investigating these effects, see Schlesinger's comprehensive analysis (1968, pp. 44-70). For a recent correlational study of transformations and comprehension, with several hundred children in grades 4 through 6 as subjects, see Fagan's summary of his dissertation (1971).

Effects of a few of these transformations are briefly summarized below.

The active vs. the passive voice.--Coleman (1965) measured the ease with which 60 college undergraduates learned active sentences and their passive counterparts and found, consistently, that actives were better retained than passives. Schlesinger (pp. 47-48) described several studies by other investigators who had found that subjects responded to active sentences faster than to their passive transformations.

Active verbs vs. their nominals.--As previously noted (see footnote on p. 36), in his study of word classes Coleman found sentences containing active verbs to be far more comprehensible than those constructions which used nominals.

Positive sentences vs. their negative forms.--Fagan's work with elementary school children showed that negatives were among the transformations most difficult for children to comprehend; he confirmed what a number of previous investigators had found (see the review in Schlesinger, pp. 45-49).

As Schlesinger concluded, after reviewing the results of over a dozen experiments,

> Sentences which are the result of a grammatical transformation will be more difficult to understand, to read, to produce and to recall than grammatically more simple sentences.
>
> . . . . . .
>
> This finding appears to be of considerable practical importance for the writer of texts and for the teacher of language. If simplification of sentence structure is the objective, one of the means of achieving this may be to dispense with transformations.
>
> A word of caution may be in order here, lest this method be used indiscriminately. Transformations often serve important linguistic functions. It has been pointed out . . . that nominalizations lack certain specific references which are bound to occur in the simpler version of the sentence, and therefore they might be preferred on certain occasions. Similarly, the passive construction permits of greater conciseness in that the actor does not have to be mentioned (as in The house was built). Occasionally, clumsiness of style can be avoided by using the passive transformation. [pp. 69, 70]

As a measure of complexity, the number and kind of grammatical transformations is a relative newcomer. In contrast, for over fifty years researchers have been using the length of a sentence as a measure of its complexity. Relatively recent measures of the sentence, which are now replacing the typical word (or syllable or letter) count, are potentially more useful both in predicting and in controlling reading ease. These newer indexes are examined at some length in the next section.

Sentence Length and Sentence Structure

The length of a sentence is at best a crude measure of its complexity. Only a little reflection suggests that often sentences can be made long or short quite arbitrarily, that any sequence of indepen-

dent clauses can be strung together with "and's" or broken up by periods without, probably, having much effect on readability. Where the choice is simply between a long sentence and a shorter statement of the same idea, we prefer the short form; undoubtedly it is easier to read. But a simple count of the number of words in a sentence is often not adequate, for it does not necessarily reflect its complexity. True, sentences of complex structure are often long ones, but investigators have found more useful measures than the number of words in a sentence.

Independent clause length.--The length of a sentence is probably not much more than a symptom of reading difficulty, at least for all but the youngest readers (Schlesinger, 1968, pp. 78-80); in contrast, clause length may very well be a cause, as suggested in a number of experiments by Coleman (1965), in which 60 undergraduates were asked to memorize and then repeat sentences prepared by the investigator. The purpose of the study was to compare several grammatical transformations with their detransformations. Clause length in the simpler versions was often shorter. As Coleman reported,

> In 52 of the detransformations, one long clause was changed into two short coordinate clauses, for example, A knowledge of the Mississippi would be helpful. → If you knew the Mississippi, it would be helpful. In 33 of the pairs, the transformation having two clauses was better retained, in 13 the opposite was true, and in 6 there was no difference . . . . Apparently a person can process content morphemes packaged into two clauses more easily than he can process the identical morphemes packaged into a single clause. Thus it seems that the advice to prefer short sentences might be better rephrased as a rule to prefer short clauses. If the clauses in a writer's composition are short, he will probably not improve readability much by emphasizing the boundaries between them

with periods and capitals. [pp. 340-341][6]

In a correlational study Bormuth (1966) found that "measures of length based upon the independent clause yield higher correlations with passage difficulty than those based upon the sentence [p. 188]." This measure suggests why (a) is better than (b):

(a) When they included this provision, they admitted the importance of the system.

(b) Their inclusion of this provision is admission of the importance of the system. [Coleman, 1965, p. 333]

The independent clause in sentence (a) is 7 words in length; in (b) it is 13 words, almost twice as long. Readability decreases as clause length increases.

Researchers appear to be in general agreement that clause length is a better predictor of reading ease than sentence length. However, this measure, too, has its limitations. For example, it does not show why sentence (c) is superior to sentence (d):

(c) The curriculum is at best . . . a design to be interpreted by teachers, for students--by teachers with varying degrees of ability and insight, for children with differing equipment in intelligence and language background. [Northrup Frye]

(d) The curriculum is at best . . . a design to be interpreted by teachers with varying degrees of ability and insight for children with different equipment in intelligence and language background. [Christensen, 1968b, p. 575]

In (c) the main clause is 33 words long; in (d) it is 29 words--a world

---

[6] What Coleman called a "coordinate clause" most of us would label an independent (or main) clause plus a dependent clause--

Independent clause: it would be helpful
Dependent clause: If you knew the Mississippi.

of difference rhetorically between these two sentences, yet an insignificant difference in independent clause length. To discriminate between these sentences we must turn to a more refined measure, one that I have not seen applied in studies of readability; it is what English teachers call the base clause rule.

Base clause length.--To understand the base clause rule it is necessary first to distinguish between two classes of modifiers which may be added to a clause: the bound and the free modifier.[7] These modifiers may be single words, any of the numerous kinds of phrases, or dependent clauses. Bound modifiers are restrictive or limiting elements that are not separated from a clause by punctuation. Free modifiers are nonrestrictive elements that are often set off by subordinating marks of punctuation (such as commas or dashes or colons).

The term base clause can now be defined as "what is left when the free modifiers are subtracted [Christensen, 1968b]." Sentence (d), above, has no free modifiers and its base clause is as long as the sentence: 29 words. Sentence (c), however, has two lengthy free modifiers, and its base is only 14 words long. These sentences may

---

[7] These terms are adapted from Christensen's work. Francis Christensen, a leader in the teaching of composition, has expanded the base clause rule into a full rhetoric program for high school and college students. The principles underlying his approach are defined in Notes Toward a New Rhetoric: Six Essays for Teachers (1967); these are further illustrated in The Christensen Rhetoric Program: The Sentence and the Paragraph (1968a).

In "The Problem of Defining a Mature Style" (1968b) Christensen compared his approach to teaching rhetoric with that of other leaders, specifically Kellogg W. Hunt and John C. Mellon. James Moffett has compared and evaluated the work of these three authorities in Chapter 5 of Teaching the Universe of Discourse (1968). There, too, Moffett includes a discussion of his own, alternative methods.

be "diagrammed" as follows, with base clause and free modifiers numbered for level of generality.

(d) 1 The curriculum is at best . . . a design to be interpreted by teachers with varying degrees of ability and insight for children with different equipment in intelligence and language background.

(c) 1 The curriculum is at best . . . a design to be interpreted by teachers, for students--

2 by teachers with varying degrees of ability and insight,

2 for children with differing equipment in intelligence and language background.

The main clause is on the highest level of generality; clause modifiers appear at descending levels in the hierarchy. (Compare R. Brown's 1958 study of abstractness levels--see pp. 33-35 of this paper.)

Following are other sentences illustrating the base clause rule, these written by below average seventh graders.

(e) 1 Last night I saw the gypsies--

2 there were black eyed girls in scarlet shawls,

2 there were men with handkerchiefs round their throat, and silver loops in their ears.

(f) 1 At home before dinner in our kitchen it is very noisy,

2 with dishes dropping,

2 pot clangs,

2 and cupboards bang shut.

(g) 1 For supper we didn't fool around on the table

1 we sat eating like smart ones,

2 used our manners,

2 talked quietly,

2 and were most obedient.[8]

Following are additional examples of the base clause rule applied, both of the sentences from an essay by Richard M. Levine which appeared in the April 1971 issue of Harper's.

(h) 1 [Adam Clayton Powell, Jr.,] probably never realized the full extent of the feeling against him,

2 partly because he spent very little time in Harlem

[2] and partly because he had always surrounded himself with a retinue of sycophants,

3 people who owed their jobs in anti-poverty agencies to him and so were generally available to chauffeur him around or keep him company during a night on the town but were not able to tell him that he was in deep trouble even if they knew.

(i) 1 This is not as surprising as it might seem,

2 for Powell and the Southerners sat on either norm of the American racial dilemma and they had always had much in common:

3 an assured constituency and the seniority it brings;

3 a career built on race and fostered by clever use of the same two basic texts,

4 the Bible and the House Rules of Order;

3 an easygoing affability and a sense of style.

Both the vocabulary and the degree of detail in these two sentences, if not their structure, make Levine's sentences unsuitable for immature

---

[8] In this sentence it is highly dubious that "and were most obedient" is on the same level of generality as the modifiers that precede it; yet the student has punctuated the last element as though it were coordinate with the others, and she has written it in grammatically parallel form.

readers. That such words as "constituency" and "sycophants" are inappropriate is clear. What may not be so obvious is that the high degree of specificity, in itself, makes this material inappropriate for children. As Bloomer (1959) hypothesized, in a study of the relation between reading difficulty and level of abstractness (or modifier load):

> The child . . . does not need precision of writing. He fills in the details from his own emotional experience. The meaning for emotional experience as it is used here is simply that reading acts as a relatively ambiguous stimulus on which a person projects his own feelings and experience. When the material becomes so precise as to demand that the person minimize his own idea substitution and react to that of the author, he becomes frustrated. The frustration at delimiting his own projection will make the readings difficult for the immature reader, though the author may use words which the reader understands fully in other contexts. If, on the other hand, there is little concentration required and some experiences can be rewarding to the individual, he will find that positive emotion builds for the subject matter and he is able to read deeper into the subject. [p. 270]

(Compare Bloomer's conclusion with the negative correlation that Coleman [(1966) 1971] and others have found between number of adjectives and reading ease.)

Viewing the elements in a sentence as members of a hierarchy is a convenient way for a writer to diagnose a wide range of strengths and weaknesses in his work, including its abstractness level. The strategy is equally useful to the reader who would improve his comprehension. Repeated practice in differentiating between general, specific, and equal groups is one of the newer approaches to teaching reading (see Sparks & Johnson, 1971). And approaches to writing such as Christensen's Rhetoric Program (1968a) can lead to improved reading skills. As Christensen wrote,

> The reader who misses the sentence pattern usually misses it because . . . he just charges ahead, reading every sentence element as something new. This type of reader must be taught to recognize the signposts that signal coordination and subordination. He can never be a good reader unless he can, almost instantly, sense when to stay at the same level of generality and when to downshift. Simultaneously, he must sense that the sentence element he is reading is now a specific detail of a general term offered previously, now a concrete illustration of some abstract principle, now some singular treatment of a plural term. The student who has not learned this can be helped to learn it--with the help of this program. [p. 27]

One final variable which may affect the reading ease of sentences should be noted: the position of the modifier.

Position of the Modifier: The Loose vs. the Periodic Sentence

The loose, or cumulative, sentence moves deductively from the whole to the part, from the head to the modifier; the periodic sentence is ordered inductively, from the part to the whole, from the modifier to the head. Whether to place a modifier (phrase or clause) before or after the base is a choice that a writer often faces. He can withhold the main idea until the end, as in this (periodic) sentence by a seventh grader:

> With wax and cloth in hand I start dusting. (The base clause is underlined.)

Or in this one, written by an adult [Lavin, 1966]:

> Then, suddenly, as a climax, and as late as possible, and only after we have supplied all the preliminary, unconnected qualifications, we state the key idea.

Of course these writers could have reversed the order, beginning with a base clause rather than ending with it. Obviously the effect

would have been far different. The question at issue here is whether one pattern is more readable than the other. Apparently there has been little research. Schlesinger (1968) lists this as one of the many variables remaining to be investigated. His discussion continues:

> Thus it has been argued that static sentences in which the main clause precedes the subordinate clause will be easier to read for the slow reader than dynamic sentences because in the former the reader gets a "meaningful concept" at an earlier stage (Wonderly no date). The same writer suggests that the head-plus-modifier construction (maison grande) may be easier for the inexperienced reader than the modifier-plus-head construction (grande maison) because in the former the transition probabilities after the first word are higher. [p. 160]

Control Through Organization of Paragraphs

As Coleman (1968) has written,

> There is not much of an experimental nature to say here. Surely most of us believe that the major determiners of readability for adults lie at this level--lie in the associations between clauses and paragraphs, in the overall organization--but psychologists have not yet refined the experimental techniques to investigate this level and linguists are not yet able to describe it. [p. 177]

In the almost complete absence of literature specifically investigating the paragraph, we return to a rationally based point of view, one that is an extension of previous discussions.

In important ways sentences and paragraphs are structurally similar. (1) The base clause is to a sentence what a topic sentence is to a paragraph. (2) The modifiers in a sentence are written as phrases and clauses; in a paragraph these modifiers appear as whole sentences. (3) Like a paragraph, a sentence can have a concluding element, as in this sentence written by a college freshman:

> Consider how Octavius manages the match between Antony and Octavia, how he proceeds against Pompey and Lepidus, how he relentlessly closes in on Antony--in short, how he is the ingenious person of the play.

Christensen (1967, 1968a) has demonstrated fully that it is possible to analyze sentences and paragraphs in identical terms. He found the relation "so close [that] the paragraph seems to be only a macro-sentence or meta-sentence [1967, p. 54]." Indeed, sometimes different marks of punctuation are the only difference between a paragraph and a single-sentence version of the same idea. More often grammatical transformations are needed to change a paragraph into a sentence.

The writer often has a choice between these units. Because ideas can be organized more neatly in a sentence, and relations between parts made clearer, he may well prefer the sentence version. But if his readers are very young, or if the grammatical transformations in the sentence are too involved for the intended audience, then he should choose the paragraph.

Other guidelines may be given:

Subordinate sequences.--Certainly children in their writing are not able to sustain a subordinate sequence for as long as an adult can, without losing interest or getting off the track. In reading, too, they may very well "skip the details." The writer of materials for children, then, should stop before he has reached the lower levels of specificity. (The same applies when writing for a mass audience; see Bloomer, 1959.)

Additionally the writer should take care to proceed one level at a time and avoid bewildering gaps. If the following sequence of sentences

by a seventh grader is short on subtlety, his meaning at least is perfectly clear:

A Stupid Stupid

(j) 1 Stupid is the most memorable dog I ever known.

2 He was so stupid he walked right into misschief every day.

3 He would pull a clean table cloth off of the table and lay on it for a bed . . . .

Coordinate sequences.--The following paragraph illustrates both within- and between-sentence coordination. It was written by a seventh grader.

Working on Saturdays

(k) 1 On Saturday mornings I go to the farm and work either in the orchard or in the fields.

2 In the fields we go to get corn and gather the corn stalks and load them in the wagon and get the tractor and ride to the scillo to make sillage.

2 In the scillo we stand around and stamp it flat and spead it out . . . .

Studies in the language development of children (for example, those by Craig, 1970; Hunt, 1965) have shown that coordination decreases as grade level increases, that coordination between main clauses (either with or without the use of conjunctions) is found several times more frequently in fourth than in twelfth grade writing. Perhaps it can be inferred from this that paragraphs exhibiting a coordinate sequence are easier to read than paragraphs based on a subordinate sequence (that is, paragraphs containing no coordinate sentences). Perhaps younger readers

in particular need the rest that staying for a time on the same level of generality can afford.

Structural signals of meaning.--(1) Although the device can be carried too far, expressing parallel ideas in parallel form undoubtedly contributes to reading ease: the form underscores the semantic equality of ideas. (2) Even more crucial to reading ease, nonparallel ideas should be stated in nonparallel form.

(1) Signaling coordinates: As a rule, the less mature the audience and/or the more unfamiliar the ideas presented, the more formal the grammatical parallelism should be between semantically equivalent ideas.[9] What is pleasing variety to a more mature reader can baffle a younger one, and so on. In patterning the coordinate sentences within a paragraph, or the coordinate elements within a single sentence, the writer should try to anticipate his readers' response, for example

---

[9] As an example, surely Joseph Wood Krutch had not intended that the following paragraph be read by non-literature people, even adult non-English majors enrolled as juniors and seniors in college; if he had, for a particular group of undergraduates the paragraph below must be considered a failure, even though the sequence is simple coordination, and even though the topic sentence had been considerably simplified to emphasize the generality under discussion.

Had the parallel ideas been more consistently parallel in form, presumably this group of readers would have comprehended Krutch's meaning, as measured by their ability to recover the Level 2 coordinates in the paragraph.

> [Do esteemed contemporary English and American writers give evidence of a "Love of Nature"?] Is there any "Love of Nature" --as distinguished from an intellectual approval of the processes of biology--in Shaw? Does T.S. Eliot find much gladness in contemplating her? Does James Joyce's apostrophe to a river count; and is Hemingway's enthusiasm for the slaughter of animals really a modern expression of that devotion to blood sports which, undoubtedly, is a rather incongruous aspect of the English race's "Love of Nature"? In America Robert Frost is almost the only poet universally recognized as of major importance in whom the loving contemplation of the natural world seems the central activity from which the poetry springs.

how they might rate a string of sentences if they were given this scale:

(2) Signaling subordinates: Again, the less mature the audience and/or the more unfamiliar the ideas presented, the clearer the structural reinforcement should be. "I just bought a Ford, a Galaxie" is perfectly clear to the reader who knows that a Galaxie is one of a number of Ford models; but if he does not know that these terms operate on different levels of generality, the identical structures "a Ford" and "a Galaxie" are not going to tell him.

Summary and Conclusion

Readability research has come a long way from the time when about all the specialists could say was "Substitute easy words for hard ones" or "Keep your sentences short." But, as has been shown, many variables at the word and sentence levels are yet to be investigated, and the paragraph and units larger than the paragraph remain virtually unexplored territory. Still, researchers have initiated systematic studies of variables that no doubt eventually will extend the full range, from choice of single words to organization of whole books. In years to come, psycholinguists no doubt will uncover many more answers to the extremely complex question, "What makes a book readable?"

Over twenty years ago, before the psychology of language (psycholinguistics) was launched as a new area of research, Dolch sought solutions to the problem of matching a writer's language with the language of his readers. Dolch's advice is still timely:

> A third plan, and one that is the most successful, starts with the writer studying the audience he is writing for. If it is the average adult, he sits in buses or restaurants or wherever people talk, and listens to their kind of language and their type of vocabulary. If the writer is writing for children of a certain grade, he sits in the grade room of various schools for several days and follows the children to the playground, listening to how they express themselves. Then the writer, when he has the "feel" of his audience, sits down to write, keeping the audience right before him in his mind's eye. He writes directly to them. [1949, p. 147]

Dolch's plan is a personal one, and it reminds us that we are to write for <u>people</u>, not abstractions of people. For no matter how firm the theoretical and practical leads that grow out of studies in the

psychology of language, the one-to-one, often subjective relation between the writer and the reader has not changed. Not all of the variables affecting reading ease can be submitted to empirical study, as psycholinguists themselves would be the first to acknowledge.

REFERENCES

Anderson, J. A report of research on comprehension in reading. Paper presented at the Third International Reading Association World Congress on Reading, Sydney, Australia, August 7-9, 1970. ERIC ED 045 323.

Aquino, M.R. The validity of the Miller-Coleman readability scale. Reading Research Quarterly, 1969, 4(3), 342-357.

Aquino, M., Mosberg, L., & Sharron, M. Reading comprehension difficulty as a function of content area and linguistic complexity. Southwest Regional Educational Laboratory, Inglewood, California, 1 August 1968. ERIC ED 057 991.

Bloomer, R.H. Level of abstraction as a function of modifier load. Journal of Educational Research, 1959, 52, 269-272.

Bormuth, J.R. Cloze as a measure of readability. In J.A. Figurel (Ed.), International Reading Association Conference Proceedings. New York: Scholastic Magazines, 1963, 8, 131-134.

Bormuth, J.R. Readability: A new approach. Reading Research Quarterly, 1966, 1(3), 79-132.

Bormuth, J.R. New developments in readability research. Elementary English, 1967, 44, 840-845.

Bormuth, J.R. The cloze readability procedure. Elementary English Journal, 1968, 45, 429-436.

Bormuth, J.R. Development of Standards of readability: Toward a rational criterion of passage performance. Final report. June 1971. ERIC ED 054 233.

Brinton, J.E., & Danielson, W.A. A factor analysis of language elements affecting readability. Journalism Quarterly, 1958, 35, 420-426.

Brown, R. Words and things. New York: Free Press of Glencoe, 1958.

Brown, W.R. Science textbook selection and the Dale-Chall formula. School Science and Mathematics, 1965, 65, 164-167.

Chall, J.S. Readability: An appraisal of research and application. Columbus: Ohio State University, 1958.

Chester, R. A critical examination of vocabulary control in view of present research dealing with initial learning tasks. Paper presented at the meeting of the American Educational Research Association, New York, February 4-7, 1971. ERIC ED 049 894.

Chomsky, N. Syntactic structures. The Hague: Mouton, 1957.

Christensen, F. Notes toward a new rhetoric: Six essays for teachers. New York: Harper & Row, 1967.

Christensen, F. The Christensen rhetoric program: The sentence and the paragraph. Teacher's manual. New York: Harper & Row, 1968. (a)

Christensen, F. The problem of defining a mature style. English Journal, 1968, 57, 572-579. (b)

Coleman, E.B. Learning of prose written in four grammatical transformations. Journal of Applied Psychology, 1965, 49, 332-341.

Coleman, E.B. Developing a technology of written instruction: Some determiners of the complexity of prose. Symposium on Verbal Learning Research and the Technology of Written Instruction, Columbia University, 1966. In E.Z. Rothkopf & P.E. Johnson (Eds.), Verbal learning research and the technology of written instruction. New York: Teachers College Press, Columbia University, 1971.

Coleman, E.B. Experimental studies of readability. Elementary English, 1968, 45, 166-178, 316-324, 333.

Coleman, E.B. Collecting a data base for a reading technology. Journal of Educational Psychology, 1970, 61 (4, Pt. 2).

Coleman, E.B., & Miller, G.R. A measure of information gained during prose learning. Reading Research Quarterly, 1968, 3(3), 369-386.

Craig, W. Investigation of syntactic-semantic relationships in the selected writings of students in grades 4-12. In D.R. Bateman & F.J. Zidonis (Eds.), A grammatico-semantic exploration of the problems of sentence formation and interpretation in the classroom. Vol. 2. Final Report, Columbus: Ohio State University, 1970. ERIC ED 048 244.

Dale, E. Can you give the public what it wants? New York: Cowles Education Corp., 1967.

Dale, E. The words we know: A national inventory. Columbus: Ohio State University, 1973.

Dale, E., & Chall, J.S. A formula for predicting readability. Educational Research Bulletin, 1948, 27, 11-20, 37-54.

Dale, E., & Chall, J.S. The concept of readability. Elementary English, 1949, 26, 19-26.

Dolch, E.W. The use of vocabulary lists. Elementary English, 1949, 26, 142-149, 177.

Fagan, W.T. Transformations and comprehension. The Reading Teacher, 1971, 25, 169-172.

Flesch, R. A new readability yardstick. Journal of Applied Psychology, 1948, 32, 221-233.

Flesch, R. The art of readable writing. New York: Harper, 1949. (a)

Flesch, R. A dissenting opinion on readability. Elementary English, 1949, 26, 332-334, 340. (b)

Froese, V. Cloze readability versus the Dale-Chall formula. Paper presented at the meeting of the International Reading Association, Atlantic City, April 19-23, 1971. ERIC ED 051 975.

Fry, E.B. A readability formula that saves time. Journal of Reading, 1968, 11, 513-516, 575-578.

Fry, E.B. The readability graph validated at primary levels. The Reading Teacher, 1969, 22, 534-538, 750.

Gorman, A.M. Recognition memory for nouns as a function of abstractness and frequency. Journal of Experimental Psychology, 1961, 61, 23-29.

Hafner, L.E. Implications of cloze. Yearbook of the National Reading Conference, 1965, 14, 151-158.

Hunt, K.W. Grammatical structures written at three grade levels. NCTE Research Report No. 3. Champaign, Ill.: National Council of Teachers of English, 1965.

Johnson, F.S. How to organize what you write. Boston: Houghton Mifflin, 1964.

Jongsma, E.R. The cloze procedure: A survey of the research. An Occasional Paper in Reading. Indiana University, 1971. ERIC ED 050 893.

Klare, G.R. The measurement of readability. Ames: Iowa State University Press, 1963.

Klare, G.R. Comments on Bormuth's Readability: A new approach. Reading Research Quarterly, 1966, 1(4), 119-125. (a)

Klare, G.R. Some empirical predictors of readability. Symposium on Verbal Learning Research and the Technology of Written Instruction, Columbia University, 1966. (b) In E.Z. Rothkopf & P.E. Johnson (Eds.), Verbal learning research and the technology of

written instruction. New York: Teachers College Press, Columbia University, 1971.

Klare, G.R. The role of word frequency in readability. Elementary English Journal, 1968, 45, 12-22.

Lavin, A.L. Writing: Unit-lessons in composition--3A. Toronto: Ginn and Co., 1966.

Lorge, I.D. Predicting reading difficulty of selections for children. Elementary English Review, 1939, 26, 229-233.

Lorge, I.D. Predicting readability. Teachers College Record, 1944, 45, 404-419. Revised in The Lorge and Flesch readability formulae: A correction. School and Society, 1948, 67, 141-142.

Lorge, I.D. Readability formulae--An evaluation. Elementary English, 1949, 26, 86-95.

Lundsteen, S.W. Sequential levels of meaning in reading. Paper presented at the meeting of the American Educational Research Association, New York City, New York, February 4-7, 1971. ERIC ED 047 897.

MacGinitie, W.H. Discussion of Professor Coleman's paper: Developing a technology of written instruction. Symposium on Verbal Learning Research and the Technology of Written Instruction, Columbia University, 1966. In E.Z. Rothkopf & P.E. Johnson (Eds.), Verbal learning research and the technology of written instruction. New York: Teachers College Press, Columbia University, 1971.

MacGinitie, W.H., & Tretiak, R. Sentence depth measures as predictors of reading difficulty. Reading Research Quarterly, 1971, 6(3), 364-377.

McCall, W.A., & Crabbs, L.M. Standard test lessons in reading. (1st, 2nd, 3rd eds.) New York: Teachers College, Columbia University, 1926, 1950, 1961.

McLaughlin, G.H. Clearing the SMOG. Journal of Reading, 1969, 13, 210-211. (a)

McLaughlin, G.H. SMOG grading--A new readability formula. Journal of Reading, 1969, 12, 639-646. (b)

Mellon, J.C. "Epilogue" to "Transformational sentence-combining: A method for enhancing the development of syntactic fluency in English composition." Champaign, Ill.: National Council of Teachers of English, 1969. ERIC ED 034 752.

Miller, G.R., & Coleman, E.B. A set of thirty-six prose passages calibrated for complexity. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 851-854.

Moffett, J. Teaching the universe of discourse. Boston: Houghton Mifflin, 1968.

Moir, H. Linguistic factors related to style and meaning in written language. Elementary English, 1970, 47, 215-222.

Paivio, A. Paired-associate learning and free recall of nouns as a function of concreteness, specificity, imagery, and meaningfulness. Psychological Reports, 1967, 20, 239-245.

Paivio, A. A factor-analytic study of word attributes and verbal learning. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 41-49.

Paivio, A. Mental imagery in associative learning and memory. Psychological Review, 1969, 76, 241-251.

Pauk, W. A practical note on readability formulas. Journal of Reading, 1969, 13, 207-210.

Peterson, J., Peters, N., & Paradis, E. Validation of the cloze procedure as a measure of readability with high school, trade school, and college populations. Paper presented at the National Reading Conference, Tampa, Florida, December 1971. ERIC ED 058 000.

Powers, R.D., Sumner, W.A., & Kearl, B.E. A recalculation of four adult readability formulas. Journal of Educational Psychology, 1958, 49, 99-105.

Rankin, E.F., Jr. The cloze procedure--A survey of research. Yearbook of the National Reading Conference, 1965, 14, 133-150.

Rankin, E.F., Jr. Grade level interpretation of cloze readability scores. Paper presented at the National Reading Conference, St. Petersburg, Florida, December 3-5, 1970. ERIC ED 046-657.

Rankin, E.F., Jr., & Culhane, J.W. One picture equals 1,000 words? Reading Improvement, 1970, 7, 37-40.

Ratekin, N.H. The adequacy of the cloze in measuring comprehension of different logical patterns. Paper presented at the National Reading Conference, Tampa, Florida, December 1971. ERIC ED 061 012.

Ruddell, R.B. The effect of the similarity of oral and written patterns of language structure on reading comprehension. Elementary English, 1965, 42, 403-410.

Schlesinger, I.M. Sentence structure and the reading process. The Hague: Mouton, 1968.

Seels, B., & Dale, E. (Compilers) Readability and reading: An annotated bibliography. (Rev. ed.) Newark; International Reading Association, 1971.

Serra, M.C. Amplifying and simplifying instructional materials: Effect on comprehension. Elementary School Journal, 1954, 55, 77-81.

Smith, W.L. The effect of transformed syntactic structures on reading. Paper presented at the conference of the International Reading Association, Anaheim, California, May 6-9, 1970. ERIC ED 044 247.

Spache, G.D. Good reading for poor readers. (tgh ed.) Champaign, Ill.: Garrard Press, 1966. (a)

Spache, G.D. The Spache readability formula. In Good reading for poor readers. (5th ed.) Champaign, Ill.: Garrard Press, 1966. (b)

Sparks, J E, & Johnson, C.E. Read right! Comprehension power. Beverly Hills, Calif.: Glencoe Press, 1971.

Stolurow, L.M., & Newman, J.R. A factional analysis of objective features of printed language presumably related to reading difficulty. Journal of Educational Research, 1959, 52, 243-251.

Strickland, R.G. The language of elementary school children: Its relationship to the language of reading textbooks and the quality of reading of selected children. Bloomington: Indiana University, 1962.

Taylor, W.L. "Cloze procedure": A new tool for measuring readability. Journalism Quarterly, 1953, 30, 415-433.

Warfel, H. Language: A science of human behavior. Cleveland: H. Allen, 1962.

Weaver, W.W., & Kingston, A.J. A factor analysis of the cloze procedure and other measures of reading and language ability. Journal of Communication, 1963, 13, 252-261.

Yngve, V.H. Computer programs for translation. Scientific American, 1962, 206, 68-76.